

PATENT
5681-73900
SUN040032

"EXPRESS MAIL" MAILING LABEL NUMBER

EL990142772US

DATE OF DEPOSIT

1/30/04

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS
BEING DEPOSITED WITH THE UNITED STATES
POSTAL SERVICE "EXPRESS MAIL POST OFFICE
TO ADDRESSEE" SERVICE UNDER 37 C.F.R. 1.10
ON THE DATE INDICATED ABOVE AND IS
ADDRESSED TO THE COMMISSIONER FOR
PATENTS, P.O. BOX 1450, ALEXANDRIA, VA
22313-1450.


Derrick Brown

DYNAMIC MEMORY THROTTLING FOR POWER AND THERMAL LIMITATIONS

By:

Chung-Hsiao R. Wu

Atty. Dkt. No.: 5681-73900

B. Noël Kivlin/MJL
Meyertons, Hood, Kivlin, Kowert & Goetzel, P.C.
P.O. Box 398
Austin, TX 78767-0398
Ph: (512) 853-8800

BACKGROUND OF THE INVENTION

Field of the Invention

5 **[0001]** This invention relates to computer systems and, more particularly, to monitoring and controlling the temperature and power consumption of a memory subsystem.

Description of the Related Art

10

[0002] As the memory density and speed of memory devices continue to increase, the power consumption and the heat dissipation associated with these devices also increase. The performance of memory devices and of the corresponding computer system may be degraded if the power consumption and/or the temperature of the memory devices are too
15 high. Therefore, temperature and power management techniques may be implemented in computer systems to reduce performance degradation due to temperature and power consumption issues.

[0003] One method that may reduce the temperature and power consumption of a
20 memory device is to control the clock rate associated with the memory device. A temperature sensor may be positioned to sense the temperature of the memory device. The clock rate may be maintained or changed depending upon whether the sensed temperature is above or below a temperature threshold. For example, the clock rate may be maintained or increased if the sensed temperature is below the temperature threshold.
25 If the sensed temperature rises above the temperature threshold, the clock rate may be decreased. By decreasing the clock rate, the temperature and power consumption of the memory device may also decrease.

[0004] Another method that may reduce the temperature of a memory device is to change the speed of a system fan depending upon a measured temperature associated with the memory device. More particularly, the speed of a system fan that cools the memory device may be increased if the sensed temperature is above a threshold. The speed of the
5 fan may be maintained or lowered if the sensed temperature is below the threshold.

SUMMARY OF THE INVENTION

[0005] Various embodiments of dynamic memory throttling for power and thermal limitations are disclosed. In one embodiment, a memory controller is coupled to a
5 memory subsystem for controlling accesses to the memory subsystem. In addition, a temperature sensor is positioned to detect a temperature associated with the memory subsystem. In this embodiment, the memory controller is configured to selectively insert one or more idle clock cycles between a first memory access and a second memory access depending upon the sensed temperature.

10

[0006] In a further embodiment, a sensor is coupled to detect a power condition associated with the memory subsystem, such as the amount of current being drawn. In this embodiment, the memory controller is configured to selectively insert one or more
15 idle clock cycles between a first memory access and a second memory access depending upon the detected power condition.

20

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG. 1 is a block diagram of one embodiment of a computer system.

5 [0008] FIG. 2 is a block diagram illustrating aspects of one embodiment of a computer system including a representation of a dynamic over-temperature control loop and a dynamic over-current control loop.

[0009] FIG. 3 is a flow diagram illustrating a process for inserting idle cycles between
10 memory accesses.

[0010] FIG. 4 is a timing diagram illustrating insertion of idle cycles between a plurality of memory accesses.

15 [0011] FIG. 5 is a flow diagram illustrating another process for inserting idle cycles between memory accesses.

[0012] FIG. 6 is a graph illustrating changes in memory bandwidth as a result of inserting idle cycles between memory accesses.

20

[0013] While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof are shown by way of example in the drawings and will herein be described in detail. It should be understood, however, that the drawings and detailed description thereto are not intended to limit the invention to the particular
25 form disclosed, but on the contrary, the intention is to cover all modifications, equivalents and alternatives falling within the spirit and scope of the present invention as defined by the appended claims.

DETAILED DESCRIPTION

[0014] Turning now to FIG. 1, a block diagram of one embodiment of a computer system 100 is shown. In the illustrated embodiment, computer system 100 is a
5 multiprocessing system including a plurality of processors 110. Processors 110 are coupled to a plurality of memory controllers 130 via a system interconnect 120. Each memory controller 130 is in turn shown coupled to a respective memory subsystem 140. In addition, computer system 100 includes a service processor 150 coupled to processors 110, memory controllers 130, and memory subsystems 140.

10 [0015] Each memory subsystem 140 is a memory medium for storing software programs and data. Memory subsystem 140 may collectively form the main memory of computer system 100 from which programs primarily execute. A suitable memory subsystem 140 may include one or more memory devices such as Dynamic Random
15 Access Memory (DRAM) devices. For example, a plurality of banks of Synchronous DRAM (SDRAM), Double Data Rate (DDR) SDRAM, or Rambus DRAM (RDRAM) devices may be suitable. The memory devices may be incorporated on one or more memory modules, such as Dual In-Line Memory Modules (DIMMs).

20 [0016] Memory controllers 130 are configured to control memory transactions between the corresponding memory subsystems 140 and processors 110. Each memory controller 130 may be located on-chip with one or more processors 110 or may be located off-chip, for example, as part of a Northbridge chipset.

25 [0017] System interconnect 120 is illustrative of any interconnect structure for coupling processors 110 to memory controllers 130. In one embodiment, system interconnect 120 may be formed by a shared bus. In other embodiments, system interconnect 120 may be formed by a point-to-point switched network.

[0018] Each processor 110 may be representative of a processor in the SPARC™ family of processors. However, it is contemplated that in other embodiments, processors 110 may be representative of other types of processors such as processors in the x86 family of processors, for example.

[0019] In the illustrated embodiment, computer system 100 is representative of a high performance server system. Service processor 150 may be provided to control various management functions associated with such a system. For example, service processor 150 may perform various administrative functions such as, for example, configuration control including software partitioning and clustering, the control of various test functionality, and/or failover and administrator alerting functionality.

[0020] As will be described in further detail below, in the depicted embodiment, service processor 150 may further be operable to implement dynamic memory throttling techniques to limit thermal and/or power conditions associated with memory subsystems 140. More particularly, in one embodiment, service processor 150 may be configured to cause a memory controller 130 to insert one or more idle cycles between a plurality of memory accesses in response to detecting that the temperature of the corresponding memory subsystem 140 is above a predetermined threshold. Service processor 150 may further be configured to cause a memory controller 130 to insert one or more idle cycles between a plurality of memory accesses in response to detecting that the current draw of the memory subsystems 140 is above a predetermined threshold. In this matter, over-temperature and/or over-current conditions may be limited.

[0021] Referring to FIG. 2, a block diagram illustrating further aspects of one embodiment of computer system 100 is shown. Components that correspond to those shown in FIG. 1 are numbered identically for simplicity and clarity. As depicted in FIG.

2, in this embodiment, service processor 150 may include an idle cycle control function 250, memory controllers 130 may include configuration registers 230, each memory subsystems 140 may include a temperature sensors 240, and DC-to-DC converter 260 may include an over-current sensor 270. Furthermore, as illustrated, a dynamic over-temperature control loop 210 is formed by a feedback path from temperature sensors 240 to service processor 150. Likewise, a dynamic over-current control loop 220 is formed by a feedback path from over-current sensor 270 to service processor 150.

[0022] In the illustrated embodiment, each memory subsystem 140 may include a temperature sensor 240 that is positioned to detect a temperature associated with the corresponding memory subsystem 140. For example, each temperature sensor 240 may be positioned in a “hot spot” of each memory subsystem 140. In one embodiment, a “hot spot” may be a location of each memory subsystem 140 that is downstream from the airflow of a system fan.

[0023] As shown in FIG. 2, computer system 100 may also include DC-to-DC converter 260 to convert a high voltage to a low voltage, which is supplied to each memory subsystem 140 to provide power to the plurality of memory subsystems 140. In addition, DC-to-DC converter 260 may include an over-current sensor 270 to detect a current draw of memory subsystems 140.

[0024] In the illustrated embodiment, service processor 150 may include idle cycle control function 250 to determine whether to program a memory controller 130 to insert one or more idle clock cycles between memory accesses. Idle cycle control function 250, in one embodiment, may be implemented in software and may include one or more recipes or algorithms to determine whether an over-temperature and/or an over-current condition may exist in a memory subsystem 140 and, if so, at what rate to insert the one or more idle clock cycles.

[0025] Furthermore, in the illustrated embodiment, memory controllers 130 may include configuration registers 230. In one embodiment, the idle cycle control function 250 of service processor 150 may program the configuration registers 230 of a memory controller 130 so that the memory controller 130 inserts the one or more idle cycles between memory accesses.

[0026] During operation, each memory controller 130 may be configured to selectively insert one or more idle clock cycles between memory accesses depending upon the sensed temperature of the associated memory subsystem 140. Referring collectively to FIG. 2 and FIG. 3, each temperature sensor 240 is operable to detect a temperature associated with each corresponding memory subsystem 140 (block 310). In this embodiment, service processor 150 may be configured to receive a control signal from each temperature sensor 240 that is indicative of the temperature of the corresponding memory subsystem 140 and determine if each sensed temperature is greater than a predetermined threshold (block 320). Specifically, idle cycle control function 250 may be used to compare each sensed temperature with the temperature threshold to determine whether there is an over-temperature condition in one or more memory subsystems 140. If the sensed temperature is below the temperature threshold, no idle cycles are inserted between memory accesses. However, if the sensed temperature is greater than the temperature threshold (i.e., over-temperature condition), service processor 150 programs the memory controller 130 to insert one or more idle cycles between a plurality of memory accesses at a particular rate (block 330). In this embodiment, idle cycle control function 250 may also be used to determine the particular rate of inserting the one or more idle cycles so that service processor 150 may program the memory controller 130 accordingly.

[0027] It is noted that in another implementation, each temperature sensor 240 may be configured to detect the temperature of the corresponding memory subsystem 140 and determine if there is an over-temperature condition. In this embodiment, if an over-temperature condition exists in a particular memory subsystem 140, the corresponding temperature sensor 240 may send an interrupt to service processor 150. The interrupt may cause service processor 150 to program the corresponding memory controller 130 to insert idle cycles between memory accesses. In some embodiment, the idle cycle control function 250 of service processor 150 may further use the detected temperature to determine the particular rate for inserting idle cycles.

[0028] Turning now to FIG. 4, timing diagrams illustrating insertion of idle cycles between a plurality of memory accesses are shown. Referring collectively to FIG. 2 and FIG. 4, graph 410 illustrates an example of a detected temperature associated with a memory subsystem 140. Graph 420 illustrates an example of memory transactions to the memory subsystem 140 from the associated memory controller 130. As depicted, idle cycles 422 are inserted between memory accesses by the corresponding memory controller 130 after the sensed temperature rises above temperature threshold 412 (Point A), and subsequent idle cycles 424 are also inserted because the sensed temperature remains above threshold 412. However, as illustrated, after the sensed temperature falls below threshold 412 (Point B), the memory controller 130 stops inserting idle cycles.

[0029] In one embodiment, service processor 150 may be configured to provide hysteresis in the implementation of dynamic control loops 210 and 220, which may prevent the temperature and current draw of a memory subsystem 140 from violating the corresponding thresholds repeatedly within a short period of time. For example, a memory controller 130 may continue to insert idle cycles between memory accesses after the sensed temperature drops below the threshold until a specified temperature is reached.

In another embodiment, the memory controller 130 may continue to insert idle cycles for a specified time after the sensed temperature drops below the threshold.

[0030] Referring again to FIG. 2, each memory controller 130 may further or
5 alternatively be configured to selectively insert one or more idle clock cycles between memory accesses depending upon the detected power condition of memory subsystems 140. Referring collectively to FIG. 2 and FIG. 5, over-current sensor 270 may be coupled to detect a power condition associated with memory subsystems 140. For example, over-current sensor 270 may detect a current draw of memory subsystems 140 (block 510). In
10 this embodiment, service processor 150 may be configured to receive a control signal from over-current sensor 270 that is indicative of the detected current draw and determine if the detected current draw is greater than a predetermined threshold (block 520). Specifically, idle cycle control function 250 may be used to compare the detected current draw with the over-current threshold to determine whether there is an over-current
15 condition in memory subsystems 140. If the detected current draw is below the over-current threshold, no idle cycles are inserted between memory accesses. However, if the detected current draw is greater than the threshold (i.e., over-current condition), service processor 150 programs memory controllers 130 to insert one or more idle cycles between a plurality of memory accesses at a particular rate (block 530). In this embodiment, idle
20 cycle control function 250 may also be used to determine the particular rate of inserting the one or more idle cycles so that service processor 150 may program the memory controllers 130 accordingly.

[0031] It is noted that in another implementation, over-current sensor 270 may be
25 configured to detect the current draw of memory subsystems 140 and determine if there is an over-current condition. In this embodiment, if an over-current condition is detected, over-current sensor 270 may send an interrupt to service processor 150. The interrupt may cause service processor 150 to program memory controllers 130 to insert idle cycles

between memory accesses. In this embodiment, service processor 150 may use the detected current draw to determine the particular rate for inserting idle cycles.

5 [0032] Furthermore, in the illustrated embodiment of FIG. 2, idle cycle control function 250, which may be included in service processor 150, may determine the rate at which memory controllers 130 insert one or more idle cycles depending upon how much the sensed temperature and/or detected current is above the corresponding threshold. For example, idle cycles may be inserted at a higher rate if the detected temperature is 10 degrees above the temperature threshold than if the detected temperature is 5 degrees above the threshold. In this embodiment, the rate of inserting idle cycles may change as the detected temperature drops; for example, the rate may be reduced. In another embodiment, the rate of inserting idle cycles may remain constant as the temperature drops; however, the rate of inserting idle cycles may be increased if the temperature rises. Furthermore, it is noted that various other methods for controlling the temperature associated with memory subsystem 140 may be implemented; for example, idle cycle control function 250 may program memory controllers 130 with a default rate at the start and then change the rate accordingly. The embodiments described above may be similarly implemented for an over-current condition.

20 [0033] Idle cycle control function 250 may also determine the rate at which memory controllers 130 insert idle cycles depending upon the type of condition detected. For example, if an over-temperature condition is detected in one or more memory subsystems 140, the corresponding one or more memory controllers 130 may slowly increase the rate of inserting idle cycles as desired. However, if an over-current condition is detected, 25 memory controllers 130 may insert idle cycles at a relatively high rate from the start. Therefore, in this embodiment, if both an over-current condition and an over-temperature condition are detected in memory subsystems 140, idle cycle control function 250 may

program memory controllers 130 to insert idle cycles at a relatively high rate, which in this case may be the rate associated with the particular over-current condition.

5 [0034] Memory controllers 130 may insert one or more idle cycles at any rate to attempt to resolve an over-temperature condition and/or an over-current condition. In one embodiment, memory controllers 130 may be configured to insert idle cycles between every n^{th} and $n^{\text{th}}+1$ successive memory access. For example, one or more idle cycles may be inserted between every 5^{th} and 6^{th} memory access. In another embodiment, service processor 150 may program memory controllers 130 to insert one or more idle cycles or
10 to change the number of idle cycles at a linear rate. For example, memory controllers 130 may first insert one idle cycle every 10 memory accesses, then one idle cycle every 8 memory accesses, and then one idle cycle every 6 memory accesses, and so on. In another example, memory controllers 130 may begin inserting one idle cycle every 10 memory accesses, then after a predetermined amount of time memory controller may
15 insert two idle cycles every 10 memory access, and then 3 idle cycles every 10 memory accesses, and so on. In yet another embodiment, service processor 150 may program memory controllers 130 to insert one or more idle cycles or to change the number of idle cycles at an exponential rate.

20 [0035] FIG. 6 shows a graph illustrating the effect of inserting idle cycles on memory bandwidth. As shown, inserting idle cycles between memory accesses may lower the memory bandwidth, which may then cause a corresponding decrease in the temperature and current draw of memory subsystems 140.

25 [0036] It is noted that in an alternative embodiment of FIG. 2, if the sensed temperature of one of the memory subsystems 140 is above the temperature threshold, idle cycle control function 250 of service processor 150 may program all of the memory controllers 130 (rather than a subset of memory controllers 130) to insert idle cycles. In

another alternative embodiment, computer system 100 may include multiple current sources each having an over-current sensor 270. In this alternative embodiment, service processor 150 may program only the one or more memory controllers 130 associated with the over-current condition to insert idle cycles.

5

[0037] It is also noted that in alternative embodiments of computer system 100, the insertion of idle cycles may be controlled by other hardware and/or software mechanisms. For example, in some embodiments, the functionality of idle cycle control function 250 may be implemented within memory controllers 130.

10

[0038] In addition, the idle cycle control functionality described above may also be employed in systems having other specific architectures. For example, the idle cycle control functionality may be employed within a single processor system and/or a system including a single memory controller and associated memory subsystem.

15

[0039] Although the embodiments above have been described in considerable detail, numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.

20